STUDENTS' DISCRIMINATION OF GERMAN CONTRASTS

AFTER 1 YEAR OF DUAL IMMERSION

by

Rachel Haynes

A thesis submitted to the faculty of
The University of Utah
in partial fulfillment of the requirements for the degree of

Master of Arts

Department of Linguistics

The University of Utah

December 2016

The University of Utah Graduate School

STATEMENT OF THESIS APPROVAL

The thesis of **Rachel Haynes**

has been approved by the following supervisory committee members:

Johanna Watzinger-Tharp	, Chair	**11 May 2016**
		Date Approved
Rachel Hayes-Harb	, Member	**11 May 2016**
		Date Approved
Fernando Rubio	, Member	**11 May 2016**
		Date Approved

and by **Edward J. Rubin** , Chair/Dean of

the Department of **Linguistics**

and by David B. Kieda, Dean of The Graduate School.

ABSTRACT

Research has demonstrated that an early age of onset may be advantageous for second language acquisition. Generally, such studies use subjects who have learned the target language through naturalistic immersion. However, few studies explore the phonological acquisition of students in immersion education programs. This study investigates whether first graders with 1 year of German dual language immersion (DLI) (experiment subjects) are able to perform significantly better than their nonimmersion peers (control subjects) on an oddity vowel discrimination task contrasting the vowels /ʊ/ and the German /ʏ/. The experiment and control groups were tested using an "odd man out" paradigm with four options: vowel A, B, or C as the odd vowel in a change trial, or X indicating that all vowels in the trial were the same. The DLI group did not perform better than the nonimmersion group; however, the nonimmersion group was more likely to select "same" while the DLI group more often selected A, B, or C. These results may indicate that DLI students with 1 year of immersion experience have begun forming a new phonological category for the German vowel /ʏ/ but are not yet able to correctly distinguish it from their more established /ʊ/ category.

TABLE OF CONTENTS

CHAPTER 1

INTRODUCTION

Immersion education began in the 1960s with the French Canadian immersion model (Gayman 2000). Programs grew in popularity in the 1980s (Fortune & Tedick 2008) and have continued to increase throughout the US. The state of Utah in particular has seen a surge in immersion programs: as of the 2015–2016 school year, 138 schools in Utah offered an immersion program in Spanish, French, Portuguese, Chinese, or German.

Immersion programs differ from traditional formal language education. In immersion programs, students acquire literacy and academic content in two languages, their native and a target language. For example, students learn content such as math, literature, or science *through* the target language. Various types of immersion education models exist, but in order to be classified as immersion, programs must teach content through the target language during a specifically delineated time and for at least 50% of the school day. Immersion programs also rely heavily on community support. These stipulations differentiate immersion programs from more traditional formal language education (Fortune & Tedick 2008).

Immersion education differs from traditional language education, but it is also different from full, naturalistic immersion. Immersion education still takes place in a classroom, and input in the target language is limited to teachers, fellow students, and perhaps some classroom aides or administration. In a naturalistic immersion setting,

learners are completely immersed in the target language, which is also the ambient language: for example, imagine a native Korean speaker moving to the United States and hearing English during nearly every interaction. Immersion education occupies a space between naturalistic immersion and traditional formal learning, and research with naturalistic or traditional language learners should not be generalized to language learners in immersion education programs. Immersion programs are distinct; research about immersion should be as well.

There are studies specifically about immersion education, but most of these deal with the social and sociolinguistic effects of immersion, the cognitive or academic benefits of immersion, or the holistic language accomplishments of immersion students. There is a dearth of studies regarding more specific aspects of linguistic acquisition, such as phonological perception or production, in immersion education students. Researchers in second language acquisition often focus their efforts on naturalistic immersion instead of immersion education. The current study aims to add to the body of second language acquisition research that deals with specific phonological acquisition of immersion education students.

This study examines the effects of language immersion on second language (L2) German students in order to answer the following research question:

- Do first graders who have completed 1 year of German dual language immersion perform significantly better on an oddity vowel categorization task identifying differences between /ʏ/ and the German /ʊ/ than their nonimmersion peers?

Data were collected about German immersion students' discrimination of German vowel

contrasts with their nonimmersion peers' discrimination of the same vowel contrasts. The experiment group population was the two Dual Language Immersion (DLI) first grade classes at West Elementary in Tooele and a non-DLI class in Salt Lake City School District. Subjects were between 6 and 7 years old. Testing consisted of an oddity discrimination task.

CHAPTER 2

REVIEW OF THE LITERATURE

History of Immersion Education

In the following section, overviews of the history of immersion education, current definitions of immersion programs, and immersion education in the US today will help clarify the ways in which immersion programs differ from traditional language programs. Immersion programs differ from full immersion in the target language environment (i.e. naturalistic second language learning) and from traditional formal language programs, in which instruction time and exposure to the target language is limited and language is treated as its own subject rather than as a mode of instruction (i.e. traditional classroom second language learning). Because immersion programs differ from naturalistic second language learning and from formal second language learning, they provide a unique opportunity for research on second language acquisition. This study seeks to take advantage of that opportunity and to add to the growing body of linguistic research on immersion programs.

Immersion education began in the 1960s with the French Canadian immersion model (Gayman 2000). Parents of English-speaking Canadian children recognized the importance of French proficiency in their community and wanted their children to become bilingual and biliterate in French in addition to their native English. These parents and local educators started the first French immersion program in Canada. The

first- and second-grade classes were taught completely in the target language, French; English was gradually introduced as a language of instruction, and by fifth grade, instruction was equally divided with 50% French and 50% English (Fortune & Tedick 2008, Lightbown & Spada 1994). The students in this program were linguistically homogenous: all were native English speakers learning French. This linguistic homogeneity is called one-way immersion; all students are traveling in one direction, with the goal of moving from English monolinguals to becoming English and French bilinguals (Fortune & Tedick 2008).

One-way immersion programs have slightly different cultural and linguistic goals than two-way bilingual immersion programs. Two-way bilingual immersion grew rapidly in the US in the 1980s as interest grew in supporting nonnative English speaking children in both maintaining their heritage language and integrating culturally and linguistically into the US school system (Fortune & Tedick 2008). Students in two-way bilingual immersion are linguistically heterogeneous: classrooms contain native speakers of both a prominent heritage language and the ambient language—for example, both native Spanish speakers and native English speakers. Classes are taught in both languages. Thus, in two-way bilingual immersion programs, students are moving in two directions to become bilingual and biliterate—native English speakers learning the target language, and speakers of the target language learning English (Fortune & Tedick 2008, Christian 1997).

Although one- and two-way immersion programs differ, both fit under the umbrella of immersion education. Fortune and Tedick (2008:9–10) stipulate that immersion programs clearly delineate class time dedicated to each language, use the

target language for instruction at least 50% of the time during elementary school years, promote additive bilingualism, employ teachers who are proficient in the instruction language, and have support from the relevant language community or communities. They also note that in an immersion classroom, language should be content driven and that language, culture, and content should be integrated. These stipulations, especially the high amount of time required, the emphasis on community support, and the necessity of content-driven teaching, distinguish immersion programs from traditional formal second language learning, in which students may have only limited exposure to the target language and are less likely to receive content-driven instruction.

In the US, the most common two-way immersion programs are for Spanish and English, but two-way immersion exists for other languages as well, including Mandarin, Korean, and French. One-way immersion programs, which do not rely on a population of target-language speakers, have expanded to include 18 languages in the US (Fortune & Tedick 2008). As of the 2014–2015 school year, 118 schools in Utah offered immersion programs in Spanish, French, Portuguese, Chinese, or German (see Utah State Office of Education). The board of education in Utah calls these immersion programs DLI, regardless of whether they are one- or two-way immersion programs.

The growing number of immersion students provides an opportunity for second language acquisition research. Students in elementary immersion programs are exposed to the target language at a young age but in an environment that differs from naturalistic second language learning and traditional classroom second language learning. The current study takes advantage of students in this unique instructional immersion environment by focusing on the German one-way immersion school at West Elementary

in Tooele, Utah.

Age of Acquisition

One possible benefit of elementary immersion programs is the young age of the learners. This section details various studies about age effects on language acquisition in naturalistic settings, then discusses the problems with generalizing apparent age effects in naturalistic immersion to traditional classroom second language learning. Finally, this section argues that immersion is different enough from both naturalistic and traditional classroom second language learning to merit its own research. Immersion may show some similar benefits to naturalistic immersion, or it could prove to be more similar to traditional classroom instruction; however, it is problematic without more research to generalize immersion programs and assume that age effects on language learning in immersion are identical to the effects of either naturalistic or traditional classroom instruction.

Intuition and casual observations show that younger learners seem to pick up second languages more easily than adults and achieve more nativelike production. However, the theories behind a "younger is better" hypothesis are hotly contested. Researchers disagree on which developmental stages are critical to achieving nativelike language production; which aspects of language acquisition are most affected by age; whether a younger age of onset (AOO) does indeed enhance language acquisition abilities; and, if younger really is better for acquiring language, *why* that is the case.

In his overview of critical period research, Scovel (2000:215) acknowledges the wide "variation among researchers on which age spans they use to divide up their subjects …" and notes that "there may be multiple critical periods at varying age levels

for different linguistic modalities." Many theories about language acquisition pinpoint puberty as the onset of declining language acquisition abilities. Johnson and Newport (1989) studied the effects of age on native Chinese and Korean speakers' acquisition of English. Their findings indicate that speakers who had learned English after age 17 (presumably after puberty) are significantly less nativelike in both morphology and syntax than speakers who had learned English before age 15 (presumably before puberty). However, their findings also show a steady decrease in nativelikeness before age 15: speakers with an AOO of 3–7 were more nativelike that those with an AOO of 8–10, who in turn were more nativelike than those with an AOO of 11–15. Johnson and Newport's findings offer some evidence for a leveling off after puberty, but they also indicate possible benefits for learning a language during early childhood (e.g. 3–4 years) as opposed to later (e.g. 8–10 years). Benefits of early childhood language acquisition are also supported by Flege and colleagues (1999), who found that as the AOO for native Korean learners of English increased, the target language grew steadily less nativelike. Their study, however, did not indicate any difference before or after puberty, but rather a steady decline from early childhood into adulthood.

There is also disagreement regarding which aspects of second language acquisition (if any) decline with an older age of acquisition. Overall, pronunciation seems to be affected by age of acquisition (Flege et al. 1995). Other studies have shown that morphosyntactic competence may also be affected by age (Johnson & Newport 1989).

Generally, studies seem to show that AOO does affect nativelikeness in language acquisition. However, *why* age seems to matter is perhaps even more controversial. The initial conception of the critical period was based on biological changes to the brain, but

Flege (1987) pointed out that there is no conclusive evidence for a biological critical period; confounding factors of age and maturation could play a part in why children seem to acquire language more readily than adults. For instance, children could lack cognitive maturity, which may help them avoid overanalyzing the language, or other maturational, educational, or social factors could contribute to declining language ability (Flege et al. 1995). However, discovering the mechanism behind age factors in language acquisition is beyond the scope of this study and appears to be peripheral to the concerns of immersion programs; that a younger age of acquisition seems to make language acquisition easier is enough to motivate and justify language immersion programs for elementary students.

The above studies about nativelikeness and the Critical Period Hypothesis suggest that an immersion program that begins when children are at a young age may have the potential to produce students with near-nativelike abilities in the target language. However, these studies on nativelikeness and the Critical Period Hypothesis are mostly gleaned from research on naturalistic immersion—that is, children and adults moving to a new environment in which the L2 is the ambient language. Of course, naturalistic immersion of this kind is quite different from immersion education, although immersion programs try to simulate naturalistic immersion in many ways. When learners are exposed to the target language in a naturalistic immersion setting, they hear the language both in and outside of school and they are likely obligated to speak it with many of their peers, even if their first language (L1) is spoken at home. Thus, the quality and quantity of input in a formal school setting, including immersion, cannot compete with the input in a naturalistic setting.

In her article “Symmetries and Asymmetries of Age Effects in Naturalistic and

Instructed L2 Learning," Carmen Muñoz (2008) discusses the dangers of applying studies of second language acquisition in a naturalistic setting to pedagogical practices and assumptions about learning in instructed settings. Muñoz (2008:578) asserts that "the differences in the amount and quality of the respective input of the two learning settings may have a significant influence on the effects that the age of initial learning has on the outcome of second language learning." In other words, without enough quality and/or quantity of input, the benefits of starting to learn a language at an earlier age may not manifest at all. Muñoz (2008:591) explains that simply lowering the age of introduction to formal language instruction will not automatically give students significantly better language acquisition the way it seems to in a naturalistic setting, because students lack the same quality and quantity of input.

Muñoz's (2008) concerns about equating naturalistic language exposure with formal language exposure are well founded; differences in quality and quantity of input do affect learners' progress in the target language. Muñoz (2008:578–79) describes formal language instruction as having

> some or all of the following features: (i) instruction is limited to 2–4 sessions of approximately 50 minutes per week; (ii) exposure to the target language during these class periods may be limited in source … quantity … and quality (there is a large variability in teachers' oral fluency and general proficiency); (iii) the target language is not the language of communication between peers; and (iv) the target language is not spoken outside the classroom.

This description, however, does not apply to immersion programs, and Muñoz (2008:579) indeed concedes that "in school immersion situations … input and use of the target language may … be limited but to a much lesser extent." In immersion programs, target language instruction ranges from 50% to 90% of students' classroom time; by definition, immersion education programs use the target language for instruction at least

50% of the time during elementary school years (Fortune & Tedick 2008). The second feature may apply somewhat to immersion programs; language exposure is limited in source in immersion programs as well as in more traditional programs, as students usually get most of their input from a teacher and perhaps a teacher's assistant. However, as mentioned, quantity is much greater in an immersion program, and the quality of input from the teachers is carefully monitored; for example, the DLI teachers in Utah are required to have at least a level of Advanced Mid on the Oral Proficiency Interview scale (see Utah State Office of Education). The third and fourth features, that the target language is not used between peers or outside the classroom, are likely true of one-way immersion programs in the early years, but as students continue in the program they may speak to peers in the target language both inside and outside the classroom. Two-way immersion programs are even more likely to see students using the target languages among peers and in the community.

Do students in immersion programs have enough exposure to the target language to benefit from exposure at an early age? Muñoz (2008:582) seems to regard this as a possibility, saying, "an early starting age produces long-term benefits when associated with greater time and massive exposure, as in immersion programmes." Immersion falls between naturalistic and traditional formal exposure to language and, as such, deserves to be treated as its own subject of research for linguists interested in second language acquisition. The present study considers immersion programs separately from naturalistic and traditional formal language learning and seeks to contribute to research about this unique type of second language learning.

Research on Immersion Education Programs

With the increase of immersion education programs, scholarship about immersion education programs has also increased. Most studies fall into one of three categories: the effects of L2 immersion on cognition and academic success (in content areas such as mathematics, social studies, or science); the effects of immersion programs on students' social or sociolinguistic identities and attitudes; and the students' linguistic development in the L2. In this section, I give an overview of studies and show that while cognitive, academic, and social benefits have been researched thoroughly, research on L2 linguistic development is relatively sparse.

Cummins (1979:223) argued against the theory that bilingualism could "cause … cognitive confusion," positing instead that bilingualism could positively influence cognitive development. Many studies explore the effects of bilingualism on cognition and show that bilingual education does not harm students' success in content subjects. Here I review three such studies.

Christian (1997) studied 3 two-way, Spanish-English elementary school immersion programs in Virginia, California, and Illinois. Using standardized tests, she assessed third-, fourth-, and fifth-grade immersion students' competence in mathematics, science, and social studies. Christian found that immersion students from all three schools progressed in these subject areas at least as well as their nonimmersion peers (with the exception of one school's fifth graders in social studies and science). Christian noted that a higher proportion of gifted students in the immersion programs may have contributed to their academic success, but the immersion programs did not seem to hinder students' academic success.

Nicolay and Poncelet (2012) noted that early bilinguals (i.e. bilingual children who learn multiple languages from their homes or communities) have been found to have cognitive advantages over monolingual children, but that the cognitive benefits of bilingualism for children in immersion programs have not been studied as frequently. To address this gap, Nicolay and Poncelet (2012) compared the cognitive development of 8-year-old native-French-speaking students in an English immersion group to that of a group of their monolingual peers. The study found that immersion students do have similar cognitive advantages to early bilinguals: both immersion students and early bilinguals are faster than their monolingual peers on "tasks assessing alerting, auditory selective attention, divided attention, and mental flexibility" (Nicolay & Poncelet 2012:597), although immersion students were not significantly better than their monolingual peers in response inhibition tasks.

Like Nicolay and Poncelet (2012), Bialystok and Barac (2011) studied cognitive benefits in immersion students. Bialystok and Barac studied two variables, level of L2 proficiency and length of time in the immersion program, in two environments: native English, Russian, and Hebrew speakers in a Hebrew immersion school, and native-English-speaking students in French immersion schools. In both studies, the two variables correlated with benefits to the students. Level of proficiency was positively correlated to performance on metalinguistic tasks, such as determining the semantic and grammatical correctness of a sentence. Length of time in the immersion program was positively correlated with performance on executive control tasks, such as matching objects by color while successfully disregarding their shape. Bilinguals' enhanced abilities in both metalinguistic and executive control tasks had been noted, but always

with early bilinguals—Bialystok and Barac showed that these cognitive benefits extended to students in immersion programs.

The work by Christian (1997), Nicolay and Poncelet (2012), and Bialystok and Barac (2011) demonstrates the general findings of studies focused on the cognitive and academic benefits for immersion students: immersion programs seem to provide students with certain cognitive advantages, such as greater executive control, and immersion programs do not seem to hinder students' performance in academic subjects that are taught through the L2.

In addition to cognitive and academic benefits, effects of immersion programs on students' social and sociolinguistic attitudes have also been extensively researched, primarily in two-way Spanish immersion programs and primarily through qualitative research methods. Most of these studies find that students' attitudes towards the target language are positive and that students are generally eager to learn and use the target language.

Potowski (2004) studied the social aspects of a two-way Spanish immersion classroom, using observational data and interviews with students to determine how often, with whom, and for what purpose Spanish was most often used. She closely observed four fifth-grade students and found that Spanish was used about 56% of the time; that girls (regardless of their L1) tended to use Spanish more; that students used Spanish more with the teacher than with peers; and that Spanish was mainly used for on-task purposes, while English use covered a wider range of topics. Potowski's research shows that there seemed to be a diglossia in which Spanish and English were used for separate purposes. However, it does not show any negative attitudes toward either language.

Morales (2012) conducted a qualitative case study on effects of two-way immersion in the school, the classroom, and the individual students. She used surveys, interviews, and observational data in two fifth-grade classrooms and found that, although native English speakers still had some cultural advantages compared to English learners, Spanish was generally accepted and normalized. The normalization of Spanish lead to L1 Spanish students succeeding at school, maintaining ties with their heritage, and integrating into the school system, while also increasing multicultural sensitivity in L1 students.

Brumen (2011) studied the motivational factors for kindergarteners in a one-way immersion program in Slovenia. She conducted interviews with the immersion students through teachers and found that students seemed motivated and that a large majority of students (96%) seemed to enjoy learning the target languages (German and English). Brumen (2011) found that for kindergarteners, external motivational factors were not very important (e.g. students were not worried about grades) but that students had an "inborn curiosity to explore the world" (Brumen 2011:725), which contributed to their motivation to learn and use the target languages.

The studies by Potowski (2004), Morales (2012), and Brumen (2011) demonstrate the general consensus that students in immersion programs seem willing and even eager to learn and use the target language. Additionally, exposure to a second language in immersion programs seems to help increase students' multicultural sensitivity and, for two-way programs, may help minority language speakers integrate into the majority language environment.

Studies on cognitive, academic, and social benefits to students generally support

the effectiveness of immersion programs. However, studies on the students' development of the target language are much rarer, and the few studies that exist do not seem to reach a clear consensus. The following section reviews several studies on specific aspects of L2 phonological acquisition and explains the consistencies and discrepancies in their findings.

Harada (2006) conducted a study on native-English-speaking children in a Japanese immersion program. He elicited the immersion students' production of both singletons and geminates in Japanese words. The students' productions were then measured acoustically and rated by native Japanese speakers for nativelikeness and accentedness. Harada found that the immersion students' production was not nativelike when compared to monolingual Japanese children—the immersion students produced longer geminates than the native speakers. However, Harada speculates that the students' *perception* may have been more nativelike than their production; the results of the study seemed to show that students could differentiate between Japanese singletons and geminates, a distinction that is not present in English. However, Harada's experiment did not test perception, so further studies are necessary to determine the extent of the students' perceptual abilities in Japanese. Harada's study leaves open the possibility that students may become nativelike or closer to nativelike in perception but not in production or that nativelike production may develop more slowly than perception.

Netelenbos (2013) cites many studies on L2 phonological research that look at acquisition of the L2 phonological system of bilinguals in a context in which the L2 is the majority language (e.g. Baker & Trofimovich 2005, Flege et al. 1999), but she explains that there are not many studies looking at the L2 system of students in immersion

programs. Netelenbos addresses this in her study of L2 phonetic development of native English speakers learning French in an immersion environment, looking specifically at the voice onset time (VOT) values of voiced and voiceless plosives to determine whether students seem to develop two distinct sound systems for their L1 and L2. Like Harada (2006), Netelenbos found that students' production of French was not nativelike but that their perceptual boundaries were similar to native French speakers' VOT boundaries.

Both Harada (2006) and Netelenbos (2013) found that immersion students' perception of the target language was similar to native speakers' perception, while the students' production was not nativelike. However, Darcy (2012) found the opposite in her study of a one-way immersion program in Germany. Looking quantitatively at both L2 vowel perception and production, Darcy studied native Turkish speakers acquiring new, difficult German contrasts. To assess students' perception, Darcy designed an oddity discrimination task, on which they performed differently from native speakers. However, the students' production of these contrasts was judged to be mostly nativelike.

The language studies on production and perception do not provide clear data, and, as Netelenbos (2013) explains, few qualitative studies on immersion students' target language development have been conducted. It is therefore difficult to establish a consensus about immersion students' L2 perception or production. The combination of young age and communicative, content-based instruction should yield high proficiency for immersion students; more quantitative data on language acquisition in immersion contexts would help determine how high immersion students' L2 proficiency actually is. Further quantitative research on specific aspects of students' language development will be beneficial to both immersion program studies and to research on second language

phonology.

L2 Perception and the German Vowel System

This study will seek to contribute to the body of quantitative research on aspects of L2 language development by exploring the sound perception of immersion students. This section will provide background for existing research on L2 perception, as well as specific perceptual difficulties for English learners of German, the target language of interest for this study. Perception is an important and complex aspect of L2 acquisition, as perception and production seem to be related when learners are acquiring a second language. Additionally, L1 categories likely contribute to L2 perception; the smaller impact of a first language phonological system on younger learners may explain why younger learners seem better able to acquire nativelikeness in production than their adult counterparts.

L2 Perception

As noted earlier, young learners seem to acquire second languages more readily than adults, especially in regard to L2 nativelike sound production (Flege et al. 1995). Jacewicz (2002) finds that perception and production are related and do not develop separately from each other, so a better understanding of L2 perception may help lead to a better understanding of nativelike production.

Many researchers agree that first-language phonology likely contributes to the perception of sounds in second languages (Best et al. 1988, Best & Tyler 2007, Flege et al. 1999). According to Best and Tyler (2007) and Mayr and Escudero (2010), L2 speech perception is related to a combination of length of residence, relative usage of the L1 and

L2, and quantity and quality of input from native L2 speakers. According to Best and Tyler (2007:20), experienced learners may also be able to "categorize and discriminate certain nonnative L2 contrasts … better than less experienced listeners, but generally less well than native L2 speakers," although the authors note that the difference between experienced and inexperienced listeners is not operationalized.

Additionally, it is well documented that infants are exceptional at differentiating all linguistic contrasts, regardless of their first language, before approximately 8 months of age, an ability that is likely influenced by lack of first language interference (Werker & Tees 2002). By approximately 12 months of age, infants' discrimination between nonnative contrasts appears to decline (Werker & Tees 2002). Although a steep decline in nonnative contrast discrimination happens during infancy, some studies seem to show that production of nonnative sounds does not level off after infancy. In other words, a younger age of exposure—even if that age of exposure occurs after infancy—seems to lead to more nativelike production of second languages (Johnson & Newport 1989). Thus, exposure to a second language in elementary school could contribute to more nativelike production, even if the children in an elementary immersion program were not exposed to the target language in infancy. However, as discussed above, exposure to the target language in a formal setting should not be equated with exposure in a naturalistic setting.

Other studies have examined which types of nonnative contrasts are easiest for adults to discriminate. According to Best's Perceptual Assimilation Model (PAM), some contrasts are easier for L2 speakers to perceive than others. Best and Tyler (2007) identify six categories of L2 contrast perception and speculate how difficult each contrast

would be for an L2 speaker to learn. First, in Two Category (TC) assimilation, two phones fit into two separate categories in the L1. Since there is a perceived difference, this contrast would be easy to learn. For example, native English speakers have no difficulty assimilating German contrasts /d/ and /t/, since this contrast exists in both languages. Second, in Single Category (SC) assimilation, both L2 phonological categories are equivalent to the same L1 phonological category; minimal pairs would be perceived as homophones. This contrast would be quite difficult to learn. One example of SC assimilation is native Japanese speakers' perception and production of English /l/ and /r/. Third, in Category Goodness (CG) assimilation, both L2 phonological categories are equivalent to the same L1 phonological category, but one is perceived as more deviant than the other. This contrast would be less salient than the contrast in TC, but since there is a perceived difference, learners could develop a new category for the deviant phone, making the contrast easier to acquire than SC assimilation. For example, native English speakers learning Arabic map both /k/ and /q/ to English /k/, but they perceive /q/ as a strange or deviant version of /k/.

The other three categories of L2 contrast perception deal with sounds that are uncategorized or nonassimilable to the L2 learner. Fourth, in Uncategorized-Categorized assimilation, learners assimilate one sound to an L1 category, but the other sound does not fit into any native category. These differences should be discriminated quite easily. Fifth, in Uncategorized-Uncategorized (UU) assimilation, learners do not assimilate new L2 categories with L1 categories. This contrast may be easy or difficult to learn, depending on phonological attributes of the phonemes in question and their relation to L1 categories. Finally, some sounds may be so deviant that learners hear them as sixth, Non-

Assimilable, or not as speech sounds at all. This category, like UU, may be either easy or difficult to learn, based on acoustic differences between the two sounds (Best & Tyler 2007:22–23).

To summarize, L2 speech perception seems to be influenced by relative usage of the L1 and L2, AOO of the L2 use, and quantity and quality of input from native L2 speakers. Additionally, perception seems to vary for different types of contrasts (Best & Tyler 2007). Further studies of L2 perception, conducted specifically on immersion students, will help contribute not only to a better understanding of the effects of immersion programs, but also to a more thorough understanding of L2 phonological systems.

In order to examine the perception of elementary school German learners, as is the goal of the current study, we must understand the German and English phonological systems. German contrasts that are difficult for English second language learners to acquire will supply the most information about English L1 systems and how they influence acquisition of German contrasts.

Polka (1993) tested adult monolingual English speakers' discrimination of several German vowel contrasts: /ʏ/, /ʊ/, /y/, /u/, /ø/, and /ɔ/. Since English does not have a phonemic distinction between high front and high back vowels that is independent of lip rounding—in English, all high front vowels (tense /i/ and lax /ɪ/) are unrounded, while all back vowels (tense /u/ and lax /ʊ/) are rounded—Polka hypothesized that monolingual English speakers may have difficulty distinguishing between front and back high vowels such as /y/ and /u/ because they are not also distinguished by lip rounding. Polka used an AXB study to measure monolingual English speakers' discrimination between vowel

pairs. Subjects heard three syllables, the first (A) and third (B) of which were different. The second syllable (X) was the same as either the first or third syllable. Subjects were asked to choose whether the second syllable more closely resembled the first syllable (A) or the third syllable (B). Polka found that English speakers discriminated sound pairs well above chance. However, the lax vowel pair discrimination, /ʏ/ and /ʊ/, was significantly lower (86.9% accurate) than the other contrasts, which were near ceiling. Polka concluded that the lax vowel contrasts fit into the CG difference assimilation from Best and Tyler's (2007) PAM: English monolingual speakers map the front /ʏ/ to the back /ʊ/, but the English speakers seem to perceive front /ʏ/ as a "less good" exemplar of that vowel category.

Jacewicz (2002) and Mayr and Escudero (2010) also examined vowel contrasts in German, finding, like Polka (1993), that the front rounded vowels prove particularly difficult for native English speakers to discriminate. Although English speakers did not usually confuse rounded and unrounded vowels with similar placement (e.g. /ɪ/ and /ʏ/), two rounded vowels with different placement, /ʊ/ and /ʏ/, did pose perception and production difficulties for native English speakers. Like Polka (1993), Jacewicz (2002) found that English speakers had particular difficulty discriminating between the lax vowels /ʊ/ and /ʏ/. The findings of Polka (1993), Jacewicz (2002), and Mayr and Escudero (2010) suggest that studying learners' ability to perceive a difference between /ʊ/ and /ʏ/ will yield valuable insight into the reorganization of the native English speakers' vowel system to accommodate new contrasts in the L2.

To understand the DLI students' perception of the German vowel contrast /ʊ/ and /ʏ/, this study will borrow heavily from Darcy and Krüger's 2012 study of perception and

production in children learning a second language. Darcy and Krüger (2012) compared the perception and production of native-Turkish-speaking 10-year-old children who began learning German between the ages of 2 and 4 with the perception and production of their native-German-speaking peers. They used an engaging, child-friendly oddity-discrimination task to test the subjects' discrimination between four German vowel contrasts. Subjects completed 96 trials in which they listened to three robots and chose the odd one out or selected the "same" option if they believed all robots said the same word.

Darcy and Krüger's 2012 study has some differences from the current study. First, Darcy and Krüger (2012) focused on both perception and production, while the current study looks only at perception (and thus borrows techniques only from the perception segment of Darcy and Krüger's study). Additionally, the target language for Darcy and Krüger's students was the ambient language (German in Berlin), while the current study's subjects are learning a foreign language (German in Tooele, Utah). Because Darcy and Krüger's research goal was to compare L2 learners to native speakers, their control group was native-German-speaking children. In the current study, the target language is a foreign language, and students are not readily exposed to the L2 outside of their school. The goal of the current study is not to compare L2 learners with native speakers, but rather to compare L2 learners with their monolingual peers in order to determine whether the perception of first-grade German immersion students differs from their peers not exposed to German. Thus, the control group for this study is not native speakers of the target language, but rather native English speakers with no exposure to German.

Despite these differences, Darcy and Krüger's 2012 study is a good model for the present study. Because Darcy and Krüger's study, like this one, investigated young L2 learners' perception of vowel contrasts, their methodology is suitable for the current study. Darcy and Krüger's odd-one-out discrimination task is engaging and simple enough for young subjects. However, the current study adjusted the number of trials from 96 to 48 because of the younger age of the subjects (ages 6 to 8 instead of 10). With this adjusted methodology, the current study aims to answer the following research question:

- Do first graders who have completed 1 school year (approximately 9 months) of German dual language immersion perform significantly better on an oddity vowel categorization task identifying differences between /ʏ/ and the German /ʊ/ than their nonimmersion peers?

This study aims to add to research about second-language learners who are in neither a fully naturalistic language setting nor in a traditional nonimmersion instructional setting, but in between: in an instructional setting that emulates aspects of naturalistic immersion. The current study will examine specific phonological aspects of immersion learning, adding to other studies about immersion programs that focus mainly on social, cognitive, or academic benefits or on more holistic language goals.

CHAPTER 3

METHODOLOGY

Subjects

In order to conduct an experiment with first-grade students as subjects, the researcher received permission from the superintendent of the two participating school districts, Tooele County School District and Salt Lake City School District. Principals and first-grade teachers of the participating classes also gave permission for the researcher to pull individual students from class for 15–30 minutes each. The teachers in these classrooms distributed permission forms to parents and students. The researcher also introduced herself in the three classes to explain the purpose of the experiment and allow subjects to become familiar with her.

Students from the two DLI first-grade classes made up the experiment group. Although 45 first graders and their parents volunteered to participate, three opted to end their sessions and did not complete the experiment, leaving 42 subjects in the experiment group. All of the experiment group first graders were between 6 and 8 years old. Their age of first exposure to German was between 5 and 7 years old (their age at the start of the 2014–2015 school year). Subjects self-reported that their primary language in the home was English. All subjects from the experiment group had very similar exposure to German. Their main source of German was their teacher, a native English speaker whose level of proficiency is rated as Advanced Mid on the Oral Proficiency Interview scale,

which is required for DLI teachers in Utah. Students also had exposure to a native-German-speaking teacher's aide, who led some class exercises and worked in small groups and individually with students for 6 weeks at the start of the school year. The Utah DLI model stipulates that students spend 50% of their class time in the German classroom and 50% with the English teacher. In preparation for this study, the researcher spent 15 hours observing Tooele's DLI classroom and can confirm that the teacher used German exclusively, which is required for all DLI programs in Utah.

The study's control group subjects were first-grade students from the two first-grade classrooms at Ensign Elementary in Salt Lake City, Utah. Ensign Elementary is not a DLI school, so none of the students had regular exposure to German during the school day. Only seven students and their parents from the first-grade class at Ensign Elementary agreed to participate in the study. The control students were the same ages, 6 to 8 years old, as their experiment group counterparts. These subjects also self-reported that English was their primary language at home.

The researcher carefully recruited subjects to form the experimental and control groups. However, since subjects were drawn from existing classrooms, placement of subjects into experimental and control groups was not randomized. With the experimental subjects coming from one classroom, teacher, and location and control subjects from another, there are multiple confounding factors that are unavoidable in this type of research. To draw any conclusions from the data, it is necessary to assume that confounding factors, such as school location, had a negligible effect on the results.

Additionally, the control group was much smaller than the experiment group. Parents in the non-DLI classroom were less willing to allow their children to participate.

The discrepancy in group sizes makes statistical comparisons difficult.

Stimuli

As discussed in the literature review, studies, including Polka (1993), Jacewicz (2002), Mayr & Escudero (2010), have shown that native English learners of German typically have difficulties differentiating between the vowels /ʏ/ and /ʊ/. Thus, the vowels of interest in this study are the lax, near-close, near-front rounded vowel /ʏ/ and the lax, near-close, near-back rounded vowel /ʊ/. These vowels differ only in fronting. To create filler tokens, we also created stimuli using the open-mid front unrounded vowel /ɛ/, which is present in both English and German.

The consonantal context /dVt/ was used to create our stimuli. Our nonwords were /dʏt/, /dʊt/, /dɛt/, and /dit/. The onset and coda were both stop consonants, /d/ and /t/. This followed the stop onset and coda pattern of stimuli in Darcy and Krüger (2012), which used /pVp/ and /kVk/ constructions. Darcy and Krüger's exact pVp and kVk constructions were not used because they created words identical or similar to common English words that first graders are likely to know: pup, kick, etc. The dVt construction created nonwords or words that first graders were less likely to know, such as "debt" or German "Dutt" (hair bun), while still being possible constructions in both German and English.

These nonwords were recorded in the carrier sentence "Sag ____ für mich" ("Say ____ for me") (see Jacewicz 2002) in order to elicit a more natural speech environment. Three native German speakers were recorded in a quiet room using a laptop recorder and the program Praat. Each speaker read each of the four stimuli sentences six times, creating six individual tokens of each sound per speaker. Of the six individual tokens for

each vowel sound, the four clearest tokens were chosen for use in the experiment. Thus, the experiment used four instances of the four nonwords from each speaker, or 48 total individual tokens. The stimuli were manually cut from the carrier sentence using the program Praat, a computer software program for phonetic speech analysis.

Procedure

The researcher pulled individual subjects out of their classes during the school day. Each subject took between 15 and 30 minutes to complete the test. Subjects were tested in quiet rooms in their schools. The experiment was hosted on SCORM Cloud and accessed online. In the DLI school in Tooele, subjects were tested in a quiet classroom used primarily to help individual students with specific needs during class time. Therefore, during test time, other students came in for help with reading or math from a tutor, so the room was not perfectly quiet. In the non-DLI school in Salt Lake City, Internet was not available for wireless devices. Subjects were tested in the main office on an office desktop computer. Like the classroom in Tooele, the office also held distractions such as parents and children coming in and out for various reasons. All subjects were accompanied by the researcher throughout the experiment.

For both the DLI group and the non-DLI group, the test was explained and presented in English. Presentation in English was necessary for the monolingual English speakers in the non-DLI group, but the presentation language for the German learners in the DLI group was carefully considered. Some studies (e.g. Antoniou et al. 2012) show that language mode may affect subjects' perception of the tokens. Delivering the test explanation in English may have primed the subjects in the current study to be in English mode before the test, and this may have affected the way they categorized the vowels in

the odd-man-out task. If the subjects were primed to listen for English vowel sounds, they may have ignored differences between /ʊ/ and /ʏ/, which are not separate phonemes in English. A German mode of presentation may have primed DLI students to listen for differences between /ʊ/ and /ʏ/. It is possible that the DLI students would have differentiated between /ʊ/ and /ʏ/ more readily had the experiment been presented in German.

However, presenting the test to the DLI group in German would have introduced different problems in the study. First, because the subjects were beginning learners of German, they may not have fully understood the task if it had been explained in German. Additionally, presenting to the DLI group in German and the control group in English would have created another variable between the groups. The potential drawbacks to both English and German presentation were carefully weighed, and English was chosen as the presentation mode that was less problematic. Thus, the test was presented in English for both groups.

The task consisted of a training portion, which contained 12 tasks, and an experiment portion, which contained 48 tasks. Three "monster" characters appeared on screen for each task. The monster characters were identical in every way except color. For every task, the monsters appeared in the same order: the purple monster on the left, the green monster in the center, and the red monster on the right. Each monster character had the same voice throughout the experiment. The interstimulus interval was 500 ms (see Darcy & Krüger 2012).

The researcher explained the purpose of the experiment in age-appropriate language. She also emphasized that subjects could ask questions any time and stop

participating at any point. The researcher then introduced the subject to the Captivate program, showing a sample image of the monsters and explaining the oddity discrimination task. Subjects were told that each monster character had a consistent voice and that each character would say a word. The subjects had to listen closely to the words to determine whether all words were the same or if one was different. If the words were all the same, the subject would click the red X at the bottom of the screen. If one character said a different word, the subject would click on the character who had said the odd word. The researcher gave examples of words, such as "doot, doot, deet," while explaining the task. The researcher then helped the subject put on headphones and made sure the subject could hear the program. Volume was set to 50% for all subjects.

Next, subjects began the training portion of the experiment. During the training portion, subjects heard 12 trials. These 12 trials contained only the vowel sounds [i], [ʊ], and [ɛ], which are all phonemes in both English and German and should be easy for monolingual English speakers to differentiate. Six of the trials were "same" trials, in which all three characters said the same word; six were "different" trials, in which character A, B, or C said an odd word. During the training portion, each vowel was the "odd" vowel twice (contrasting once with each of the other two vowels) and each character said the "odd" word twice. The order of the trials was random. See Table 1 for a list of the trials and combinations. Subjects were not able to repeat audio for the training or testing portions of the experiment.

During the training portion, subjects received feedback from the program indicating whether they had answered correctly or incorrectly. The researcher paid attention during the trial period to answer any questions subjects might have and to help

them click the “continue” and “submit” buttons. The researcher also recorded manually whether responses were correct or incorrect during the trial period. In order for their testing portion responses to be considered valid for analysis, subjects were required to answer at least 75% of the trial portion questions correctly (see Escudero et al. 2008). This requirement was intended to ensure that participants understood the task and were able to correctly use the testing software. All subjects fulfilled this requirement and were thus asked to continue to the testing phase of the experiment.

After completing the training portion of the experiment, subjects were shown a slide in the Captivate program saying “Good work!” The researcher then explained that they would continue to the next phase of the experiment. The researcher explained that the next part of the experiment would be slightly different from the first part in several ways: subjects would not receive feedback for their answers; the next trial would play as soon as the answer was selected; and the sounds would be different and more difficult, so subjects should pay close attention. After hearing this introduction to the testing portion, subjects pressed continue and began the testing portion of the experiment.

The testing portion of the experiment consisted of 48 trials total. All answers for the experiment portion were recorded. Twelve of the 48 trials were fillers that did not test subjects’ ability to distinguish between /ʊ/ and /ʏ/. These filler trials consisted only of vowels /ɛ/ and /i/, which are phonemes in English as well as German and should have been easy for subjects to differentiate (see Table 2 for details of the filler trials). The filler trials were randomly interspersed with the 36 testing trials. The filler trials were used as a check for the researcher to see whether subjects were actively listening and focusing during the testing portion: if subjects did not answer more than nine of the 12 filler trials

correctly (75% correct), the researcher assumed that they either did not understand or were not engaged in the task. Seven of the subjects from the DLI class failed to answer nine or more filler trials correctly; their answers were therefore removed from the results.

The 36 testing trials tested the subjects' ability to differentiate between /ʊ/ and /ʏ/. Eighteen of these trials were classified as "same" trials, in which the three monsters all said the same nonword, either /dʊt/ (for nine trials) or /dʏt/ (for nine trials). For these trials, the correct answer was X. The other 18 trials were "different" trials, in which one of the three monsters said a different nonword than the other two. For the "different" trials, /dʊt/ was the odd word out nine times: three times from Monster A, three from Monster B, and three from Monster C. The same was true for the nonword /dʏt/ (see Table 3 for details).

The researcher sat near the subjects while they completed the testing portion of the experiment. The researcher was on hand to answer any of the subjects' questions and to observe any technical difficulties or clarify confusing aspects of the test program. Because of the subjects' young age, the researcher also stayed close to encourage subjects to finish the experiment. If subjects became bored, the researcher would encourage them once to finish the experiment, but she would also remind them that they were allowed to leave at any time. If subjects requested to leave, the researcher sent them back to their classroom. Subjects were rewarded with a plastic alphabet ruler for completing the experiment. When subjects completed the experiment, the researcher asked them what languages they spoke at home with their parents. Subjects then returned to their classrooms.

Table 1. Training Trials

Trial #	Monster A vowel	Monster B vowel	Monster C vowel	Correct answer	Odd vowel
1	[i]	[i]	[i]	X	–
2	[ʊ]	[ʊ]	[ɛ]	C	[ɛ]
3	[ɛ]	[ɛ]	[ɛ]	X	–
4	[i]	[i]	[i]	X	–
5	[ʊ]	[ɛ]	[ɛ]	A	[ʊ]
6	[ɛ]	[ɛ]	[ɛ]	X	–
7	[ʊ]	[i]	[ʊ]	B	[i]
8	[ʊ]	[ʊ]	[ʊ]	X	–
9	[ɛ]	[i]	[i]	A	[ɛ]
10	[ɛ]	[ɛ]	[i]	C	[i]
11	[i]	[ʊ]	[i]	B	[ʊ]
12	[ʊ]	[ʊ]	[ʊ]	X	–

Table 2. Filler Trials

Number of occurrences	Monster A vowel	Monster B vowel	Monster C vowel	Correct answer	Odd vowel
3	[ɛ]	[ɛ]	[ɛ]	X	–
3	[i]	[i]	[i]	X	–
1	[i]	[ɛ]	[ɛ]	A	[i]
1	[ɛ]	[i]	[ɛ]	B	[i]
1	[ɛ]	[ɛ]	[i]	C	[i]
1	[ɛ]	[i]	[i]	A	[ɛ]
1	[i]	[ɛ]	[i]	B	[ɛ]
1	[i]	[i]	[ɛ]	C	[ɛ]

Table 3. Experiment Trials

Number of occurrences	Monster A vowel	Monster B vowel	Monster C vowel	Correct answer	Odd vowel
9	[ʊ]	[ʊ]	[ʊ]	X	–
9	[ʏ]	[ʏ]	[ʏ]	X	–
3	[ʊ]	[ʏ]	[ʏ]	A	[ʊ]
3	[ʏ]	[ʊ]	[ʏ]	B	[ʊ]
3	[ʏ]	[ʏ]	[ʊ]	C	[ʊ]
3	[ʏ]	[ʊ]	[ʊ]	A	[ʏ]
3	[ʊ]	[ʏ]	[ʊ]	B	[ʏ]
3	[ʊ]	[ʊ]	[ʏ]	C	[ʏ]

CHAPTER 4

RESULTS

Analysis

One clear problem with the data is the disparity of group sizes: the DLI group consisted of 35 qualifying individuals (i.e. individuals who correctly answered 75% of the filler trials), while the control group consisted of only seven individuals. This difference in size makes the groups difficult to compare and limits the conclusions we can draw from statistical tests. It is important to acknowledge the limitations that the small control group size presents, but we will nevertheless present and analyze the available data.

The collected data from the DLI group and the control group were analyzed using t-tests and d-prime analysis. Two t-tests were run on the raw data. Then, the d-prime was calculated for each subject, and two more t-tests were run on the d-prime averages for each group. The t-tests compared how often the control group and the experiment group answered questions correctly and how often each group answered "change" (A, B, or C) or "same" (X). D-prime controlled for response bias. While the raw data yielded nearly significant differences, the differences between the groups' d-prime scores were insignificant. The following section will explain these tests.

Using the first t-test, we compared how often each of the two groups chose the correct answer to the 36 experiment trials: A, B, C, or X. This directly addressed the

research question: Do first graders who have completed 1 year of German dual language immersion perform significantly better on an oddity vowel categorization task identifying differences between /ʏ/ and the German /ʊ/ than their nonimmersion peers?

If DLI students could correctly choose A, B, C, or X during the experiment significantly more often than their non-DLI counterparts, we could assume they were "better" at identifying differences between the /ʏ/ and /ʊ/ sounds.

The results of this first t-test were surprising. While the difference between the DLI and the control groups was nearly significant, it was significant in favor of the control group. The DLI group answered a mean of 13 questions correctly out of 36 questions, and the non-DLI group answered a mean of 16.5 questions correctly out of 36 questions. The p-value of this comparison was 0.0554, just missing significance.

These results raise the question of why the control group would perform better than their DLI counterparts on a vowel categorization task. One possible explanation for these results is that the non-DLI group more frequently selected X, or "same," when completing the experiment. Because "same" was the correct response for 18 of the 36 trials, a higher rate of guessing "same" could have improved the non-DLI students' scores in the first trial. Hypothetically, the non-DLI students may have had a tendency to ignore or overlook differences between the /ʏ/ and /ʊ/ sounds, which are not phonemes in English. They may have therefore guessed more frequently that all the sounds were the same. In this case, they would have answered more of the 18 "different" trials incorrectly, but they would have answered more of the 18 "same" trials correctly.

Furthermore, it is possible that DLI students were more attuned to the presence of phonemes /ʏ/ and /ʊ/ in German after 1 school year of exposure to the German language

but that they could not yet accurately identify an odd vowel in the oddity discrimination task. In this case, they may have been more prone to guessing A, B, or C more frequently but may not have been able to accurately identify which token was different. Thus, they may have missed more of the 18 “same” tasks without correctly identifying A, B, or C in the “different” tasks. In this case, the non-DLI students would have had an advantage in the “same” trials while the DLI students did not have an advantage in the “different” trials.

To test this theory, a second t-test was run, this time comparing how frequently each group selected X, or “same,” during the 36 experiment trials, as opposed to selecting A, B, or C (“different”). The DLI group answered “same” (X) an average of 20 times out of 36 trials. The non-DLI group answered “same” an average of 26 times. The p-value for this test was 0.0742, also nearing but missing significance. The tendency, though not significant, was for the non-DLI group to select “same” more often than the DLI group.

While the control group more frequently answered “same” than the DLI students, a t-test on the raw data is insufficient to determine whether the higher number of “same” responses is truly the cause of the higher accuracy of correct answers shown in the first t-test. In other words, there is not necessarily a correlation of “same” answers with the control group’s correct answers. To determine whether or not the "same" answers correlate with higher accuracy of the control group's answers, a d-prime analysis was run to control for response bias.

D-prime analysis tested signal detection effects to see whether either group reliably detected “same” trials in contrast to “different” trials. The d-prime score was calculated by comparing each subject’s hits to false alarms. A hit resulted from students

correctly selecting “same.” A false alarm resulted from students selecting “same” when the correct answer to the trial was actually one of the “different” options (either Monster A, B, or C). If an individual’s d-prime score is close to zero, that individual is inaccurate at detecting a signal.

The mean d-prime score for the DLI group was 0.24; the mean d-prime score for the non-DLI group was 0.33. Both of these scores are close to zero, thus showing that neither group was able to accurately detect when a trial was “same” versus “different.” Furthermore, a t-test between the two groups’ d-prime mean scores showed that there was no significant difference between the two groups’ mean scores. Due to the large difference in group size, a second t-test was run on the d-prime scores, comparing seven randomly selected DLI students to the seven control students. The t-test for this sample also showed that the means of the two groups were not significantly different.

Our analysis shows that the control group was not more accurate at selecting “same” than the DLI group; based on their low d-prime scores, both groups of students were unable to detect a difference between the “same” and “different” trials. Therefore, the control group’s higher selection of “same” answers does not explain why the control group would have higher accuracy than the DLI group.

Apparently, neither the control nor the DLI group was able to detect a difference between the “same” and “different” trials, as shown by the d-prime analysis. However, the control group did tend to select “same” more frequently than the DLI group. There could be many reasons for this difference, especially with such a small sample of control students. One possible explanation is that the DLI students were more willing to listen for and guess at differences between the sounds /ʏ/ and /ʊ/ than their counterparts. Although

this does not explain why their performance in accuracy was poorer than their control counterparts, it may still show that they were more attuned to the presence of phonemes /ʏ/ and /ʊ/ in German after 1 school year of exposure to the German language.

Discussion

This study sought to answer the research question

- Do first graders who have completed 1 school year (approximately 9 months) of German dual language immersion perform significantly better on an oddity vowel categorization task identifying differences between /ʏ/ and the German /ʊ/ than their nonimmersion peers?

The results of the study show, in short, that the answer is no: the DLI first graders in this study did not perform better on the categorization task than their nonimmersion peers. Surprisingly, the nonimmersion students performed almost significantly better than the DLI students. However, various factors in the way the research was conducted complicate the results of this study: the findings are unexpected and somewhat questionable, and this study is best regarded as a pilot study to prompt future research about DLI.

It is difficult to draw conclusions from this study, largely because of concerns with the subjects and the stimuli used in the experiment. The most obvious concern is the size of the control group compared to that of the experiment group. The experiment group, which consisted of DLI students, had 35 viable subjects; the control group had only seven. Schools, teachers, and parents of non-DLI students were less willing to participate in the study, which resulted in the discrepancy of group sizes.

Recruiting non-DLI subjects proved much more difficult than recruiting DLI

subjects. Of the three first-grade classes at West Elementary, two participated in the German program; the third class used only English throughout the school year. Originally, this researcher planned to use students from the non-DLI class as control subjects. Six total students' parents agreed to allow their children to participate. However, three of those students had special needs. The high percentage of students with special needs within the control group would have confounded any differences in results between the groups. Thus, control subjects were found through a different elementary school class, at Ensign Elementary in Salt Lake City. However, parents, students, and teachers at Ensign Elementary were not invested in German, DLI, or studies exploring the effects of DLI. Thus, only a small segment of students and their parents were willing to participate in this study. The discrepancy in size between the DLI and non-DLI groups makes statistical comparisons difficult; with only seven students in the control group, differences between groups are likely to be magnified or missed entirely.

Additionally, due to the nature of the experiment, it was impossible to randomly assign subjects to either the experiment or the control group. With the experimental subjects coming from one classroom, teacher, and location and control subjects from another, multiple confounding factors were unavoidable in this research. To draw any conclusions from the data, we would have to assume that confounding factors, such as school location, had a negligible effect on the results.

In addition to concerns with the number and distribution of subjects, there are several problems with the stimuli used in the study. Although great care was taken in creating stimuli nonwords and carrier sentences, using native German speakers, and recording tokens in a quiet room, the stimuli were problematic. Firstly, the speakers

themselves are similar as far as their residence in the US, age, and sex. All three speakers have lived in the US for many years and have therefore been far removed from German as the dominant language. Additionally, all three speakers are women over the age of 60. Speakers who differ in age, sex, and residence would better control for possible bias in pronunciation and would present a wider range of German speakers.

The use of native German speakers presents further problems due to their difference from the DLI teacher in the classroom. The DLI teacher is not a native speaker of German, and it is possible that she did not differentiate between /ʏ/ and /ʊ/ in a nativelike way. While the students in the DLI classroom may have been exposed to German input from other sources, such as video and audio materials, their nonnative teacher was the primary source of German language input for the DLI subjects. Thus, the current experiment may have tested a contrast to which the students had not been adequately exposed.

Another concern with the stimuli comes from the recordings themselves: the laptop recording equipment and overemphasized pronunciation of the nonwords combined to make the recordings of the nonwords less than ideal. All speakers emphasized the nonword within the carrier sentence "Sag ____ für mich," but one speaker in particular emphasized the nonword by pronouncing it with a high pitch. Thus, when the nonword was cut from the sentence, one monster character—Monster A—repeatedly said its word in a higher tone than the others. This may have drawn attention to Monster A. Additionally, the quality of the laptop recordings was less clear than it would have been in a sound booth. The researcher tried to mitigate this problem by recording each speaker's nonwords six times and then choosing the clearest four

recordings from those six; however, this selection process was quite subjective. Ideally, the stimuli would have been more uniform in pitch and more clearly recorded, although it could be argued that the unclear recordings are a better sample of actual speech: most conversations include background noise and some unclear pronunciation.

The concerns with the subjects and stimuli in this study make the results questionable. However, the patterns in the data are still worth discussing. The results of the first t-test, which showed that the DLI students were *less* likely than their non-DLI peers to correctly select the odd one out, were unexpected. Although the researcher hypothesized that the non-DLI group's higher performance was due to their answering X more often and more accurately, the d-prime analysis disproves this explanation: non-DLI students did answer X more often, but their X answers were not more accurate than the DLI group's X answers. Thus, the non-DLI students' performance must be attributed to something else.

While the results of the first t-test were unexpected, the results of the second t-test, which compared how often each group answered X compared to A, B, or C, were in line with the researcher's expectations. For this test, A, B, and C were classified together as a "different" answer, while X was classified as a "same" answer. The second t-test showed that the non-DLI group was more likely to answer X, or "same," than the DLI group. There are many possible explanations for this, but one interesting possibility is that the DLI group may have been more likely to select "different" (A, B, or C) because they were starting to listen for differences between /ʏ/ and the /ʊ/. In other words, the non-DLI subjects may be mapping both sounds to the English /ʊ/, the only category where the /ʏ/ sound could fit, and so selected "same" frequently. The DLI subjects, on the

other hand, may have already begun to establish a new, separate /ʏ/ category, and so selected "different" more frequently. If this was the case, the data show that DLI subjects were not accurate when selecting "different." Rather, the DLI subjects were simply more able to recognize the possibility that a difference between /ʏ/ and /ʊ/ could exist.

The DLI subjects may have been less likely to select X than the non-DLI subjects for reasons other than a restructuring of their vowel categories. It is possible that the non-DLI subjects used a strategy of selecting X more frequently because they recognized that X was the correct answer more often than A, B, or C. This strategy could be eliminated by removing the option of X during the experiment: subjects would be forced to choose between A, B, or C. This type of design could show whether one group of students is better at correctly selecting the odd man out. However, the forced-choice design would also obscure potentially important findings. In the early stages of acquiring a new vowel contrast, starting to select "different" instead of defaulting to "same" may be an important step. If the non-DLI group is indeed more likely to default to X or "same," a forced-choice design that eliminated the X option would obscure this tendency.

The DLI students' more frequent selection of "same" during the study could have interesting implications, but because of the concerns with subjects and stimuli, conclusions should be drawn cautiously. This study is best regarded as a pilot study: studies similar to this could reveal more information regarding the development of immersion students' language acquisition. The following section will outline three different types of potential future research that could build on the findings of the current study: first, a similar study conducted with more subjects and clearer stimuli; second, a study comparing DLI students to native speakers instead of non-target-language speakers;

third, various longitudinal versions of the study.

Because of the concerns with the stimuli and subjects in the current study, possible future research includes a replication of this study with the same research question but improved execution. Tokens should be created with better sound quality. Discrepancy in experiment (DLI) and control (non-DLI) group sizes must be avoided, and, ideally, future studies would include groups of equal and significant size. A replication study is a natural continuation of the current study: the data collected from a well-conducted replication study could more reliably answer the research question and could provide more dependable data for analysis. Results from a replication study could shed light on patterns that may be emerging from the results of the current study; for example, a replication study may again show that DLI subjects tend to choose "different" answers in a discrimination task more readily than their non-DLI peers, or it may reveal that there is no such pattern.

A replication study focusing on phonemes from other target languages is another possible variation for future research. The growing popularity of immersion is not limited to German: in Utah alone, Spanish, French, Portuguese, and Chinese immersion programs are already in place. Varying languages in similar programs are a ripe field for linguistic research. A replication study that tested students' discrimination of phonemes from a language other than German (e.g. the contrast between nasal and nonnasal vowels in Portuguese) could help researchers learn how students acquire a certain contrast, comparing DLI subjects to non-DLI subjects just as the current study does. A replication study could also allow researchers to compare studies between languages. Immersion programs are relatively controlled environments, so first-grade DLI students in Utah

learning Portuguese have similar quality and quantity of input to first-grade DLI students in Utah learning German. Studies could compare the patterns of acquisition of various contrasts in multiple languages.

Another possible direction for future research in this area is to compare DLI students to native-speaking peers instead of to non-target-language-speaking peers. For example, a study like this could compare first-grade German DLI students in the US to first-grade native–German-speaking students in Germany. This would be more similar to Darcy and Krüger's 2012 study, which compared native Turkish speakers learning German to native German speakers. This type of study would establish the learners' nativelikeness instead of looking for differences between two groups.

A third possibility for future research would be a longitudinal study of a group of DLI students, or a study that compares early DLI learners to students who have been in the DLI program for longer periods of time. A longitudinal study could follow a group of DLI students from first grade and track their improvement in vowel discrimination over time, throughout their elementary school DLI experience or through junior high or high school. A variation of this study could be conducted with different groups of students in an established DLI program: first-grade DLI students could be compared to DLI students of different ages. Studies such as these could help answer the question of how students establish categories over time in immersion programs. Studies conducted over a longer period of time may help clarify the data from the current experiment, answering the question of whether a propensity to select "different" in odd–man-out categories is the beginning of a better ability to discriminate between new vowel categories.

The current study yielded some unexpected and interesting results, but with the

concerns in stimuli and subjects, it is premature to draw any definite conclusions from these data. However, it is useful to consider this a pilot study for future experiments in the field of immersion education. There is plenty of room for new research regarding immersion education: immersion education's substantial differences from both naturalistic language acquisition and formal instructional language learning make it problematic to generalize existing research from these fields. A better understanding of how students acquire a language in an immersion setting can enrich understanding of how languages are learned, and further research may guide instruction and curriculum development, ultimately benefiting language pedagogy.

REFERENCES

ANTONIOU, MARK; MICHAEL D. TYLER; and CATHERINE T. BEST. 2012. Two ways to listen: Do L2-dominant bilinguals perceive stop voicing according to language mode? *Journal of Phonetics* 40.582–94.

BAKER, WENDY, and PAVEL TROFIMOVICH. 2005. Interaction of native- and second-language vowel system(s) in early and late bilinguals. *Language and Speech* 48.1–27.

BEST, CATHERINE T.; GERALD W. MCROBERTS; and NOMATHEMBA M. SITHOLE. 1988. Examination of perceptual reorganization for nonnative speech contrasts: Zulu click discrimination by English-speaking adults and infants. *Journal of Experimental Psychology: Human Perception and Performance* 14.345–60.

BEST, CATHERINE T., and MICHAEL D. TYLER. 2007. Nonnative and second-language speech perception: Commonalities and complementarities. *Second language speech learning: The role of language experience in speech perception and production,* ed. by Murray J. Munro and Ocke-Schwen Bohn, 13–34. Amsterdam: John Benjamins.

BIALYSTOK, ELLEN, and RALUCA BARAC. 2011. Emerging bilingualism: Dissociating advantages for metalinguistic awareness and executive control. *Cognition* 122.67–73.

BRUMEN, MIHAELA. 2011. The perception of and motivation for foreign language learning in pre-school. *Early Child Development and Care* 181.717–32.

CANALE, MICHAEL, and MERRILL SWAIN. 1980. Theoretical bases of communicative approaches to second language teaching and testing. *Applied Linguistics* 1.1–47.

CHRISTIAN, DONNA; CHRISTOPHER L. MONTONE; KATHRYN J. LINDHOLM; and ISOLDA CARRANZA. 1997. *Profiles in two-way immersion education.* Washington, D.C.: Center for Applied Linguistics.

CUMMINS, JAMES. 1979. Linguistic interdependence and the educational development of bilingual children. *Review of Educational Research* 49.222–51.

CUMMINS, JIM. 1980. The entry and exit fallacy in bilingual education. *NABE Journal* 4.25–59.

DARCY, ISABELLE, and FRANZISKA KRÜGER. 2012. Vowel perception and production in Turkish children acquiring L2 German. *Journal of Phonetics* 40.568–81.

ESCUDERO, PAOLA; RACHEL HAYES-HARB; and HOLGER MITTERER. 2008. Novel second-language words and asymmetric lexical access. *Journal of Phonetics* 36.345–60.

FLEGE, JAMES EMIL. 1987. The production of "new" and "similar" phones in a foreign language: Evidence for the effect of equivalence classification. *Journal of Phonetics* 15.47–65.

FLEGE, JAMES EMIL; MURRAY J. MUNRO; and IAN R. A. MACKAY. 1995. Factors affecting strength of perceived foreign accent in a second language. *The Journal of the Acoustical Society of America* 97.3125–134.

FLEGE, JAMES EMIL; GRACE H. YENI-KOMSHIAN; and SERENA LIU. 1999. Age constraints on second-language acquisition. *Journal of Memory and Language* 41.78–104.

FORTUNE, TARA WILLIAMS, and DIANE J. TEDICK. 2008. One-way, two-way and indigenous immersion: A call for cross-fertilization. *Pathways to multilingualism: Evolving perspectives on immersion education,* ed. by Tara Williams Fortune and Diane J. Tedick, 3–21. New York: Cromwell Press.

FORTUNE, TARA WILLIAMS; DIANE J. TEDICK; and CONSTANCE L. WALKER. 2008. Integrated language and content teaching: Insights from the immersion classroom. *Pathways to multilingualism: Evolving perspectives on immersion education,* ed. by Tara Williams Fortune and Diane J. Tedick, 71–96. New York: Cromwell Press.

GAYMAN, STACY MOIRA. 2000. *Understanding language use and social interaction in a French/English two-way immersion classroom.* Philadelphia, PA: University of Pennsylvania dissertation.

GREGG, KEVIN R. 1984. Krashen's monitor and Occam's razor. *Applied Linguistics* 5.79–100.

HARADA, TETSUO. 2006. The acquisition of single geminate stops by English-speaking children in a Japanese immersion program. *Studies in Second Language Acquisition* 28.601–32.

JACEWICZ, EWA. 2002. The perception-production relationship in the acquisition of second language vowel contrasts. *Journal of Language and Linguistics* 1.314–37.

JOHNSON, JACQUELINE S., and ELISSA L. NEWPORT. 1989. Critical period effects in second language learning: The influence of maturational state on the acquisition of English as a second language. *Cognitive Psychology* 21.60–99.

KRASHEN, STEPHEN D. (1982). Second language acquisition theory. *Principles and practice in second language acquisition,* ed. by Stephen D. Krashen, 9–56. Oxford, UK: Pergamon Press.

LIGHTBOWN, PATSY M., and NINA SPADA. 1994. An innovative program for primary ESL students in Quebec. *TESOL Quarterly* 28.563–79.

MAYR, ROBERT, and PAOLA ESCUDERO. 2010. Explaining individual variation in L2 perception: Rounded vowels in English learners of German. *Bilingualism: Language and Cognition* 13.279–97.

MORALES, PAOLA ZITLALI. 2012. *Examining the role of a dual language immersion program's language ideologies in scaffolding positive learner identities in dual language learners.* Los Angeles, CA: UCLA dissertation.

MUÑOZ, CARMEN. 2008. Symmetries and asymmetries of age effects in naturalistic and instructed L2 learning. *Applied Linguistics* 29.578–96.

NETELENBOS, NICOLE. 2013. *The development of voice onset time in French immersion children.* Lethbridge, AB: University of Lethbridge thesis.

NICOLAY, ANNE-CATHERINE, and MARTINE PONCELET. 2012. Cognitive advantage in children enrolled in a second-language immersion elementary school program for three years. *Bilingualism: Language and Cognition* 16.597–607.

POLKA, LINDA. 1993. Linguistic influences in adult perception of non-native vowel contrasts. *The Journal of the Acoustical Society of America* 97.1286–96.

POTOWSKI, KIM. 2010. Student Spanish use and investment in a dual immersion classroom: Implications for second language acquisition and heritage language maintenance. *The Modern Language Journal* 88.75–101.

SCOVEL, THOMAS. 2000. A critical review of the critical period research. *Annual Review of Applied Linguistics* 20.213–23.

SNOW, MARGUERITE ANN; MYRIAM MET; and FRED GENESEE. 1989. A conceptual framework for the integration of language and content in second/foreign language instruction. *TESOL Quarterly* 23.201–17.

SWAIN, MERRILL. 1993. The output hypothesis: Just speaking and writing aren't enough. *The Canadian Modern Language Review* 50.158–64.

UTAH STATE OFFICE OF EDUCATION. *Utah's statewide dual language immersion initiative.* Online: http://www.schools.utah.gov/CURR/dualimmersion/Home/ImmersionInitiative.aspx

WERKER, JANET F., and RICHARD C. TEES. 2002. Cross-language speech perception: Evidence for perceptual reorganization during the first year of life. *Infant Behavior & Development* 25.121–33.

WESELY, PAMELA M. 2009. *The language learning motivation of early adolescent French and Spanish elementary immersion program graduates.* Minneapolis, MN: University of Minnesota dissertation.